

METHOD AND DEVICE FOR VOICE SEPARATION OF COMPOUND VOICE DATA, METHOD AND DEVICE FOR SPECIFYING SPEAKER, COMPUTER PROGRAM, AND RECORDING MEDIUM

Publication number: JP2003005790 (A)

Publication date: 2003-01-08

Inventor(s): YAMAMOTO TAKAYOSHI +

Applicant(s): YAMAMOTO TAKAYOSHI; URATA TAKAYUKI +

Classification:

- International: G10L11/00; G10L15/00; G10L15/02; G10L15/20; G10L17/00; G10L21/02; G10L11/00; G10L15/00; G10L17/00; G10L21/00; (IPC1-7). G10L11/00; G10L15/02; G10L15/20; G10L17/00; G10L21/02

- European:

Application number: JP20010191289 20010625

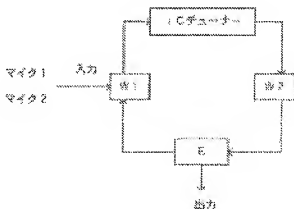
Priority number(s): JP20010191289 20010625

Also published as:

JP3364487 (B2)

Abstract of JP 2003005790 (A)

PROBLEM TO BE SOLVED: To provide a method and a device for separating compound voice data where voice data of several speakers mixedly exist into the voice of every speaker and to provide a method and a device for accurately and quickly specifying the speaker of each separated voice data. SOLUTION: The method for separating compound voice data where voice data of several speakers mixedly exist into the voice data of every speaker has a step (1) where correlation elimination processing is performed to eliminate correlation between the compound voice data and a step (2) where independent component separation processing is performed to separate data subjected to the correlation elimination processing into independent components.



(10) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2003-5790

{P2003-5790A}

(43) 公開日 平成15年1月8日(2003.1.8)

| (51) Int. Cl. | 識別記号 | F I | テーム(参考) |
|------------------------------------|-------|---------|------------------------|
| G 1 0 L | 17/00 | G 1 0 L | 3/00 5 4 5 A 5 D 0 1 5 |
| | 11/00 | | 9/00 A |
| | 15/02 | | 9/02 A |
| | 15/20 | | 9/08 A |
| | 21/02 | | 9/02 3 0 1 A |
| 審査請求 有 請求項の数21 O L (全 32 頁) 最終頁に続く | | | |

(21) 出願番号 特願2001-191289(P2001-191289)

(22) 出願日 平成13年6月25日(2001.6.25)

(71) 出願人 500262511

山本 隆義

広島市西区已斐上2丁目54-20

(71) 出願人 501253822

浦田 隆之

広島県広島市安佐南区高取南1丁目3-8

(72) 発明者 山本 隆義

広島県広島市西区已斐上2丁目54-20

(74) 代理人 100071283

弁理士 一色 健輔 (外3名)

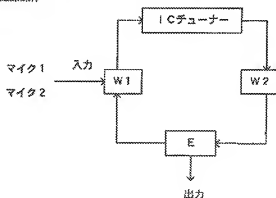
Fターム(参考) 5B015 AB03 CC04 CC05

(54) 【発明の名称】 複合音声データの音声分離方法、発音者特定方法、複合音声データの音声分離装置、発音者特定装置、コンピュータプログラム、及び、記録媒体

(57) 【要約】

【課題】 複数の発音者の音声データが混在する混在音声データを、発音者毎の音声に分離する方法及び装置、さらに分離された各音声データの発音者を特定することを、正確にかつ高確に行うことができる方法及び装置を提供する。

【解決手段】 複数発音者の音声データが混在している混在音声データを、発音者毎の音声データに分離する音声データ分離方法において、(1) 前記混在音声データを互いに無相関化するための無相関化処理を行うステップと、(2) 前記無相関化処理の行われたデータを独立成分に分離するための独立成分分離処理を行うステップとを有する。



【特許請求の範囲】

【請求項1】 複数発音者の音声データが混在している混在音声データを、発音者毎の音声データに分離する音声データ分離方法において、(1)前記混在音声データを互いに無相関化するための無相関化処理を行うステップと、(2)前記無相関化処理の行われたデータを独立成分に分離するための独立成分分離処理を行うステップと、を有することを特徴とする音声分離方法。

【請求項2】 請求項1に記載の音声分離方法において、前記独立成分分離の行われたデータの分離性が不十分な場合には、分離性が十分になるまで、前記独立成分分離処理の行われたデータについて、前記無相関化処理及び前記独立成分分離処理を繰り返し行うことを特徴とする音声分離方法。

【請求項3】 請求項1又は請求項2に記載の音声分離方法において、前記独立成分分離処理として、非ガウス性のデータを独立成分に分離するための非ガウス性独立成分分離処理と、非定常性のデータを独立成分に分離するための非定常性独立成分分離処理と、有色性のデータを独立成分に分離するための有色性独立成分分離処理とを準備し、データの性質により、前記非ガウス性独立成分分離処理、前記非定常性独立成分分離処理、及び、前記有色性独立成分分離処理のうちのいずれかの処理を行うことを特徴とする音声分離方法。

【請求項4】 請求項3に記載の音声分離方法において、

最初に行われる独立成分分離処理は、非ガウス性のデータを独立成分に分離するための非ガウス性独立成分分離処理であることを特徴とする音声分離方法。

【請求項5】 請求項1乃至請求項4に記載の音声分離方法において、前記無相関化処理は、少なくとも主成分分析及び因子分析を行うことを特徴とする音声分離方法。

【請求項6】 複数発音者の音声データが混在している混在音声データを、発音者毎の音声データに分離し、該発音者毎の音声データにつき発音者を特定する発音者特定方法において、(1)請求項1乃至請求項5のいずれかに記載の音声分離方法により、複数発音者の音声データが混在している混在音声データを、発音者毎の音声データに分離するステップと、(2)発音者毎に該発音者を特定するための特定パラメータを準備するステップと、(3)分離された前記発音者毎の音声データにつき、前記特定パラメータを参照して、発音者を特定するステップと、を有することを特徴とする発音者特定方法。

【請求項7】 請求項6に記載の発音者特定方法において、前記特定パラメータは、発音者が母音を発音した際のホ

ルマント周波数であり、

分離された前記発音者毎の音声データにつき、ホルマント周波数を求め、求められたホルマント周波数に關して、前記特定パラメータとしてのホルマント周波数を参照して、発音者を特定することとを特徴とする発音者特定方法。

【請求項8】 請求項7に記載の発音者特定方法において、前記特定パラメータは、発音者が母音を発音した際の第1ホルマント周波数及び第2ホルマント周波数であり、分離された前記発音者毎の音声データにつき、第1ホルマント周波数及び第2ホルマント周波数を求め、求められた第1ホルマント周波数及び第2ホルマント周波数に關して、前記特定パラメータとしての第1ホルマント周波数及び第2ホルマント周波数を参照して、発音者を特定することとを特徴とする発音者特定方法。

【請求項9】 請求項6乃至請求項8のいずれかに記載の発音者特定方法において、分離された前記発音者毎の音声データにつき、前記特定パラメータを参照して発音者を特定するステップにて発音者を特定できなかった場合には、該音声データから複数の時点のホルマント周波数を求め、求められた複数の時点のホルマント周波数に關して、前記特定パラメータとしての複数の時点のホルマント周波数を参照して、発音者を特定することとを特徴とする発音者特定方法。

【請求項10】 請求項6乃至請求項9のいずれかに記載の発音者特定方法において、分離された前記発音者毎の音声データにつき、前記特定パラメータを参照して発音者を特定するステップにて発音者を特定できなかった場合には、該音声データから有声音データを分離し、該有声音データにつき、ホルマント周波数を求め、求められたホルマント周波数に關して、前記特定パラメータとしてのホルマント周波数を参照して、発音者を特定することとを特徴とする発音者特定方法。

【請求項11】 請求項10に記載の発音者特定方法において、分離された前記発音者毎の音声データにつき、前記特定パラメータを参照して発音者を特定するステップにて発音者を特定できなかった場合には、該音声データから有声音データを分離し、該有声音データにつき、第1ホルマント周波数及び第2ホルマント周波数を求め、求められた第1ホルマント周波数及び第2ホルマント周波数に關して、前記特定パラメータとしての第1ホルマント周波数及び第2ホルマント周波数を参照して、発音者を特定することとを特徴とする発音者特定方法。

【請求項12】 請求項10または請求項11に記載の発音者特定方法において、

分離された前記発音者毎の音声データにつき、前記特定パラメータを参照して発音者を特定するステップにて発音者を特定できなかった場合には、

該有音声データにつき、複数の時点のホルマント周波数を求め、求められた複数時点のホルマント周波数に関して、前記特定パラメータとしての複数時点のホルマント周波数を参照して、発音者を特定することを特徴とする発音者特定方法。

【請求項 13】 請求項 10又は請求項 11に記載の発音者特定方法において、

前記音声データから前記有音声データを分離する際に、該音声データに対して独立成分に分離するための独立成分分離処理が行われることを特徴とする発音者特定方法。

【請求項 14】 複数の発音者の音声データが混在している混在音声データから、議事録を作成する議事録作成方法において、

請求項 6乃至請求項 13のいずれかに記載の発音者特定方法により、分離された前記発音者毎の音声データにつき、発音者を特定するステップと、

特定された発音者と、該発音者の発言とを対応付けて記録媒体に出力することにより、議事録を作成するステップと、を有することを特徴とする議事録作成方法。

【請求項 15】 複数の発音者の音声データが混在している混在音声データを、発音者毎の音声データに分離する音声データ分離装置において、

前記混在音声データを互いに無相関化するために無相関化処理を行い、

前記無相関化処理の行われたデータを独立成分に分離するために独立成分分離処理を行うことを特徴とする音声分離装置。

【請求項 16】 請求項 15に記載の音声分離装置において、

前記独立成分分離の行われたデータの分離性が不十分な場合には、分離性が十分になるまで、前記独立成分分離処理の行われたデータについて、前記無相関化処理及び前記独立成分分離処理を繰り返し行うことを特徴とする音声分離装置。

【請求項 17】 請求項 15又は請求項 16に記載の音声分離装置において、

データの性質により、前記独立成分分離処理として、非ガウス性のデータを独立成分に分離するための非ガウス性独立成分分離処理、非定常性のデータを独立成分に分離するための非定常性独立成分分離処理、有色性のデータを独立成分に分離するための有色性独立成分分離処理、のうちのいずれかの処理を行うことを特徴とする音声分離装置。

【請求項 18】 請求項 17に記載の音声分離装置において、

最初に行われる独立成分分離処理は、非ガウス性のデー

タを独立成分に分離するための非ガウス性独立成分分離処理であることを特徴とする音声分離装置。

【請求項 19】 請求項 15乃至請求項 18に記載の音声分離装置において、前記無相関化処理は、少なくとも主成分分析及び因子分析を行うことを特徴とする音声分離装置。

【請求項 20】 複数の発音者の音声データが混在している混在音声データを、発音者毎の音声データに分離し、該発音者毎の音声データにつき発音者を特定する発音者特定装置において、

請求項 15乃至請求項 19のいずれかに記載の音声分離装置により、複数の発音者の音声データが混在している混在音声データを、発音者毎の音声データに分離し、分離された前記発音者毎の音声データにつき、発音者毎に該発音者を特定するための特定パラメータを参照して発音者を特定することを特徴とする発音者特定装置。

【請求項 21】 請求項 20に記載の発音者特定装置において、

前記特定パラメータは、発音者が母音を発音した際のホルマント周波数であり、

分離された前記発音者毎の音声データにつき、ホルマント周波数を求め、求められたホルマント周波数に関して、前記特定パラメータとしてのホルマント周波数を参照して、発音者を特定することを特徴とする発音者特定装置。

【請求項 22】 請求項 21に記載の発音者特定装置において、

前記特定パラメータは、発音者が母音を発音した際の第 1ホルマント周波数及び第 2ホルマント周波数であり、分離された前記発音者毎の音声データにつき、第 1ホルマント周波数及び第 2ホルマント周波数を求め、求められた第 1ホルマント周波数及び第 2ホルマント周波数に関して、前記特定パラメータとしての第 1ホルマント周波数及び第 2ホルマント周波数を参照して、発音者を特定することを特徴とする発音者特定装置。

【請求項 23】 請求項 20乃至請求項 22のいずれかに記載の発音者特定装置において、

分離された前記発音者毎の音声データにつき、前記特定パラメータを参照して発音者を特定できなかった場合には、

該音声データから複数の時点のホルマント周波数を求め、求められた複数時点のホルマント周波数に関して、前記特定パラメータとしての複数時点のホルマント周波数を参照して、発音者を特定することを特徴とする発音者特定装置。

【請求項 24】 請求項 20乃至請求項 23のいずれかに記載の発音者特定装置において、

分離された前記発音者毎の音声データにつき、前記特定パラメータを参照して発音者を特定できなかった場合には、

該音声データから有声音データを分離し、該有声音データにつき、ホルマント周波数を求め、求められたホルマント周波数に關して、前記特定パラメータとしてのホルマント周波数を参照して、発音者を特定することと特徴とする発音者特定装置。

【請求項25】 請求項24に記載の発音者特定装置において、

分離された前記発音者毎の音声データにつき、前記特定パラメータを参照して発音者を特定できなかった場合には、

該音声データから有声音データを分離し、該有声音データにつき、第1ホルマント周波数及び第2ホルマント周波数を求め、求められた第1ホルマント周波数及び第2ホルマント周波数に關して、前記特定パラメータとしての第1ホルマント周波数及び第2ホルマント周波数を参照して、発音者を特定することと特徴とする発音者特定装置。

【請求項26】 請求項24または請求項25に記載の発音者特定装置において、

分離された前記発音者毎の音声データにつき、前記特定パラメータを参照して発音者を特定できなかった場合には、

該有声音データにつき、複数の時点のホルマント周波数を求め、求められた複数時点のホルマント周波数に關して、前記特定パラメータとしての複数時点のホルマント周波数を参照して、発音者を特定することと特徴とする発音者特定装置。

【請求項27】 請求項24又は請求項25に記載の発音者特定装置において、

前記音声データから前記有声音データを分離する際に、該音声データに対して独立成分に分離するための独立成分分離処理が行われることを特徴とする発音者特定装置。

【請求項28】 複数発音者の音声データが混在している混在音声データから、議事録を作成する議事録作成装置において、

請求項2乃至請求項27のいずれかに記載の発音者特定装置により、分離された前記発音者毎の音声データにつき、発音者を特定し、

特定された発音者と、該発音者の発言とを対応付けて記録媒体に出力することにより、議事録を作成することを特徴とする議事録作成装置。

【請求項29】 請求項1乃至請求項5のいずれかに記載の音声分離方法を音声分離装置に実行させるためのコンピュータプログラム。

【請求項30】 請求項6乃至請求項13のいずれかに記載の発音者特定方法を発音者特定装置に実行させるためのコンピュータプログラム。

【請求項31】 請求項29又は請求項30に記載のコンピュータプログラムを記録したコンピュータ読み取り

可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、複数発音者の複合音声データの音声とを分離する方法、分離したそれぞれの音声データの発音者を特定する方法、複数発音者の複合音声データの音声とを分離する装置、分離したそれぞれの音声データの発音者を特定する装置、コンピュータプログラム、及び、記録媒体に関する。

10 【0002】

【従来の技術】 複数の発音者の音声とが混合されて記録されている、音声記録媒体中の複合音声データを、発音者毎に正確に分離する技術が切望されている。具体的には、複合音声データを、音声の入力と同時に進行的に発音者毎に分離し特定することで、会議の議事録作成を自動的に行うことのできるような技術が切望されている。

【0003】 従来、長時間にわたる会議の議事録を作成するには、各種の音声記録機器に記録した会議の音声データを、議事録作成担当者から全て聞きおし、要約するなどして議事録を作成していた。この作業は、音声記録機器の再生と一時停止を何度も繰り返すつづける必要があり、手間と時間がかかる。

【0004】 また、もう一つの問題は、発音者の特定が困難であることである。本人が会議に出席した担当者ならまだしも、そうでない担当者が議事録を作成するのは、どの音声かどの発音者によるものなのかを判断するのは非常に困難なことであった。

【0005】 従来、混合音声データからの音声分離、発音者特定に関する技術は幾つか存在しているが、1本のマイクに複数人の音声やノイズが混合されて入力される場合でも分離、特定を正確に行い、さらに、複合音声の入力と同時に進行的に高速度な分離・特定処理を行うことは、時間的に連続な音素データのセグメンテーション、及び調音結合の点で非常に難しい課題であった。

【0006】 特開2001-27895には、複数の信号源からの音響信号を分離し、所望の信号を合成出力するための信号分離方法が記載されている。この発明は、解析対象の混合音声・音響信号に対し時間・周波数解析を行い、周波数成分の倍音構成を導る。倍音周波数成分のうち、立上がり時間及び立下り時間の少なくとも一方が共通であるか否かで、同一信号源からの周波数成分であるかどうかを判定する。その周波数成分を抽出・再構成することにより、単一信号源からの信号を分離する。

【0007】 この発明は、混合された信号の相関性や独立性といった事項を考慮していないので、同じ周波数帯域に属する混合信号、あるいは同時帯域に存在する混合信号を分離することは困難である。

【0008】 また、特開2000-97758に記載された音源信号推定装置では、複数の音源信号がそれぞれ混在して複数のチャンネルを介して入力されたときに、

各音源信号が混合係数ベクトルと内積演算されて他の音源信号に加算される混合過程モデルに基づき、混合係数ベクトルに対応する分離係数ベクトルを逐次修正しながら求め、この分離係数ベクトルを用いて音源信号の推定、分離を行う(ICAの手法)に依り、分離係数ベクトルの逐次修正に用いる修正ベクトルを正規化する音声信号とそれ以外の信号が相互に混在している信号からそれぞれの信号を推定し、分離する際し、それぞれの信号パワー変動による推定、分離への影響を軽減することができる、さらに、収束係数を大きくすることができることから安定かつ高速の信号分離が可能となる、とされている。

【0009】この発明は、独立成分解析(ICA)をベースとして分離係数ベクトルを逐次修正しながら行うので、信号パワーの変動影響を軽減でき、高速分離を実現するものであるが、様々な信号源からの音源信号はお互いに独立性を保持しているとは限らない。一般に、たとえ独立した信号源からの音源信号であっても混合されると相関性を有してしまっていることが多いが、その点が考慮されていない。

【0010】また、特開平9-258788には、基本周波数の近接した混合音声を選択的に区別分離し、音源の数に制限されず、高品質の分離音声を得ることを目的とした音声分離方法および装置が記載されている。この発明では、入力音響信号中に含まれる音声信号の有声音部分と無声音部分以外の有声音部分には有声音の音源方向の情報を加味しながら個別に抽出し、抽出された有声音部分を複数の有声音に分化して有声音の群として抽出し、音声信号の無声音部分は入力音響信号から有声音部分を減算して抽出した残差から各有声音の群の無声音に相当する音響信号の成分として抽出し、各別に抽出された有声音の群に無声音を補充して音声信号を抽出することによって上記目的を実現する。

【0011】この発明は、音源方位の情報を抽出する音源定位部を有しているが、同じ方向から異なる音声が発生した場合分岐は分離が困難となる。また、複数の発音者が同じ母音、あるいは有声音を発したときはそれらの分離が困難であると思われる。

【0012】

【発明が解決しようとする課題】以上のような従来技術が有する種々の問題点を解決すべく、本発明は、複数の発音者の音声データが混在する混在音声データを、発音者毎の音声に分離する方法及び装置、さらに分離された各音声データの発音者を特定することを、正確にかつ高速に行うことができる方法及び装置の提供を主たる目的とする。

【0013】

【課題を解決するための手段】上記の課題を解決するために、本出願に係る第1の発明は、複数発音者の音声データが混在している混在音声データを、発音者毎の音声

データに分離する音声データ分離方法において、(1)前記混在音声データを互いに無相関化するための無相関化処理を行うステップと、(2)前記無相関化処理の行われたデータを独立成分に分離するための独立成分分離処理を行うステップとを有することを特徴とする音声分離方法である。このような第1の発明によれば、入力される混在音声データ(生データ)に含まれる各音声データの相関性、及び独立性の両性質をともに考慮し、複数の音声データや混入する雑音などの有する相関性や独立性が、時間的・空間的に変動する場合でも、発音者毎の音声に正確に分離することができる。

【0014】また、本出願に係る第2の発明は、第1の発明である音声分離方法において、前記独立成分分離の行われたデータの分離性が不十分な場合には、分離性が十分になるまで、前記独立成分分離処理の行われたデータについて、前記無相関化処理及び前記独立成分分離処理を繰り返すことを特徴とする音声分離方法である。このような第2の発明によれば、混在音声データを音源毎の音声データに充分に分離させることができる。

【0015】また、本出願に係る第3の発明は、第1又は第2の発明である音声分離方法において、前記独立成分分離処理として、非ガウス性のデータを独立成分に分離するための非ガウス性独立成分分離処理と、非定常性のデータを独立成分に分離するための非定常性独立成分分離処理と、有色性のデータを独立成分に分離するための有色性独立成分分離処理とを準備し、データの性質により、前記非ガウス性独立成分分離処理、前記非定常性独立成分分離処理、及び、前記有色性独立成分分離処理のうちのいずれかの処理を行うことを特徴とする音声分離方法である。このような第3の発明によれば、無相関化処理の行われたデータの性質に応じて最適な独立成分分離処理を行うことができるから、混在音声データを音源毎の音声データにより効果的に分離させることができる。

【0016】また、本出願に係る第4の発明は、第3の発明である音声分離方法において、最初に行われる独立成分分離処理は、非ガウス性のデータを独立成分に分離するための非ガウス性独立成分分離処理であることを特徴とする音声分離方法である。非ガウス性独立成分分離処理は他の独立成分分離処理方法に比べてその前処理としての無相関化処理の影響を受けやすいから、このような第4の発明によれば、最初に非ガウス性独立成分分離処理を行うことにより、無相関化処理がうまく実行されかどうかを、該無相関化処理に引き続く非ガウス性独立成分分離処理によって効果的に評価することが可能となる。

【0017】また、本出願に係る第5の発明は、第1乃至第4の発明である音声分離方法において、前記無相関化処理は、少なくとも主成分分析及び因子分析を行うことを特徴とする音声分離方法である。このような第5の

発明によれば、各主成分の寄与率を求めて累積寄与率が所定のしきい値を超えるところの成分数を次数とすることなどにより、採用する主成分データの数(次数)を決定した上で、効果的に無相関化処理を行うことが可能となる。

【0018】また、本出願に係る第6の発明は、複数発音者の音声データが混在している混在音声データと、発音者毎の音声データに分離し、該発音者毎の音声データにつき発音者を特定する発音者特定方法において、

(1) 第1乃至第5のいずれかの発明の音声分離方法により、複数発音者の音声データが混在している混在音声データを、発音者毎の音声データに分離するステップと、
(2) 発音者毎に該発音者を特定するための特定パラメータを準備するステップと、(3) 分離された前記発音者毎の音声データにつき、前記特定パラメータを参照して、発音者を特定するステップとを有することを特徴とする発音者特定方法である。このような第6の発明によれば、例えば、会議の録音データなどに記録された、複数発音者の音声や雑音などが含まれた混在音声データを音源ごとに分離し、各分離された音声データの発音者を特定することによって、例えば、自動的に会議記録データの作成などを行うことができる。

【0019】また、本出願に係る第7の発明は、第6の発明である発音者特定方法において、前記特定パラメータは、発音者が母音を発音した際のホルマント周波数であり、分離された前記発音者毎の音声データにつき、ホルマント周波数を求め、求められたホルマント周波数に関して、前記特定パラメータとしてのホルマント周波数を参照して、発音者を特定することを特徴とする発音者特定方法である。このような第7の発明によれば、フーリエ変換などの容易な処理で抽出できる特徴量であるホルマント周波数を用いて、各分離された音声データの発音者を特定を容易に行うことができる。

【0020】また、本出願に係る第8の発明は、第7の発明である発音者特定方法において、前記特定パラメータは、発音者が母音を発音した際の第1ホルマント周波数及び第2ホルマント周波数であり、分離された前記発音者毎の音声データにつき、第1ホルマント周波数及び第2ホルマント周波数を求め、求められた第1ホルマント周波数及び第2ホルマント周波数に関して、前記特定パラメータとしての第1ホルマント周波数及び第2ホルマント周波数を参照して、発音者を特定することを特徴とする発音者特定方法である。このような第8の発明によれば、第1と第2のスペクトルピークである2つのホルマント周波数を用いて発音者の特定を行うことによって、容易に、かつより正確に特定を行うことができる。

【0021】また、本出願に係る第9の発明は、第6の発明乃至第8の発明のいずれかに記載の発音者特定方法において、分離された前記発音者毎の音声データにつき、前記特定パラメータを参照して発音者を特定するス

テップにて発音者を特定できなかった場合には、該音声データから複数の時点のホルマント周波数を求め、求められた複数時点のホルマント周波数に関して、前記特定パラメータとしての複数時点のホルマント周波数を参照して、発音者を特定することを特徴とする発音者特定方法である。このような第9の発明によれば、ある音声の発音者を特定する上での特徴量であるホルマント周波数の、時間的変動をも考慮することにより、より正確に発音者の特定を行うことができる。

【0022】また、本出願に係る第10の発明は、第6の発明乃至第9の発明のいずれかに記載の発音者特定方法において、分離された前記発音者毎の音声データにつき、前記特定パラメータを参照して発音者を特定するステップにて発音者を特定できなかった場合には、該音声データから有声音データを分離し、該有声音データにつき、ホルマント周波数を求め、求められたホルマント周波数に関して、前記特定パラメータとしてのホルマント周波数を参照して、発音者を特定することを特徴とする発音者特定方法である。ホルマント周波数による発音者の特定は、有声音、特に母音の識別に有効であるので、このような第10の発明によれば、無声音を含む様々な音声をもより正確に識別することができる。ここで、無相関化処理及び独立成分分離処理がなされる前の音声データが複数人の音声が入混在しているデータであるのに対して、無相関化処理及び独立成分分離処理という二つの処理によって分離された分離音声データは、ある一人の音声が入混在されたデータとなっている。よって、このような二つの処理によって分離された分離音声データからは有声音を高い精度で抽出することができる。

【0023】また、本出願に係る第11の発明は、第10の発明の発音者特定方法において、分離された前記発音者毎の音声データにつき、前記特定パラメータを参照して発音者を特定するステップにて発音者を特定できなかった場合には、該音声データから有声音データを分離し、該有声音データにつき、第1ホルマント周波数及び第2ホルマント周波数を求め、求められた第1ホルマント周波数及び第2ホルマント周波数に関して、前記特定パラメータとしての第1ホルマント周波数及び第2ホルマント周波数を参照して、発音者を特定することを特徴とする発音者特定方法である。このような第11の発明によれば、分離された有声音データに対して、第1と第2のスペクトルピークである2つのホルマント周波数を用いて発音者の特定を行うことによって、より正確に特定を行うことができる。

【0024】また、本出願に係る第12の発明は、第10の発明または第11の発明の発音者特定方法において、分離された前記発音者毎の音声データにつき、前記特定パラメータを参照して発音者を特定するステップにて発音者を特定できなかった場合には、該有声音データにつき、複数の時点のホルマント周波数を求め、求めら

11

れた複数時点のホルマント周波数に關して、前記特定パラメータとしての複数時点のホルマント周波数を参照して、発音者を特定することを特徴とする発音者特定方法である。このような第12の発明によれば、分離された有声音データに対して、発音者特定上の特長量であるホルマント周波数の時間的変動をも考慮することにより、より正確に発音者の特定を行うことができる。

【0025】また、本出願に係る第13の発明は、第10の発明又は第11の発明の発音者特定方法において、前記音声データから前記有声音データを分離する際に、該音声データに対して独立成分に分離するための独立成分分離処理が行われることを特徴とする発音者特定方法である。有声音は声帯の振動を伴うものなので、このような第13の発明によれば、音声データに独立成分分離処理をかけることによって、声帯の振動を伴わない無声音と声帯の振動を伴う有声音とを容易に分離することが可能となる。

【0026】また、本出願に係る第14の発明は、複数発音者の音声データが混在している混在音声データから、議事録を作成する議事録作成方法において、第6の発明乃至第13のいずれかの発明の発音者特定方法により、分離された前記発音者毎の音声データにつき、発音者を特定するステップと、特定された発音者と、該発音者の発音とを対応付けて記録媒体に出力することにより、議事録を作成するステップとを有することを特徴とする議事録作成方法である。このような第14の発明によれば、発音者の特定が自動的に正確に行われるため、長時間にわたる会議の議事録作成を自動的に行うことができ便利である。

【0027】また、本出願に係る第15の発明は、複数発音者の音声データが混在している混在音声データを、発音者毎の音声データに分離する音声データ分離装置において、前記混在音声データを互いに無相関化するために無相関化処理を行い、前記無相関化処理の行われたデータを独立成分に分離するために独立成分分離処理を行うことを特徴とする音声分離装置である。このような第15の発明によれば、入力される混在音声データ（生データ）に含まれる各音声データの相関性、及び独立性の両性質をとともに考慮し、複数の音声データや混入する雑音などの有する相関性及び独立性が、時間的・空間的に変動する場合でも、発音者毎の音声に正確に分離することが可能な音声分離装置を実現できる。また、本出願に係る第16の発明は、第15の発明である音声分離装置において、前記独立成分分離の行われたデータの分離性が不十分な場合には、分離性が十分になるまで、前記独立成分分離処理の行われたデータについて、前記無相関化処理及び前記独立成分分離処理を繰り返す行うことを特徴とする音声分離装置である。このような第16の発明によれば、混在音声データを音源毎の音声データに十分に分離させることの可能な音声分離装置を実現できる。

12

【0028】また、本出願に係る第17の発明は、第15又は第16の発明である音声分離装置において、データの性質により、前記独立成分分離処理として、非ガウス性のデータを独立成分に分離するための非ガウス性独立成分分離処理、非定常性のデータを独立成分に分離するための非定常性独立成分分離処理、有色性のデータを独立成分に分離するための有色性独立成分分離処理、のうちのいずれかの処理を行うことを特徴とする音声分離装置である。このような第17の発明によれば、無相関化処理の行われたデータの性質に応じて最適な独立成分分離処理を行うことができるから、混在音声データを音源毎の音声データにより効果的に分離させることの可能な音声分離装置を実現できる。

【0029】また、本出願に係る第18の発明は、第17の発明である音声分離装置において、最初に行われる独立成分分離処理は、非ガウス性のデータを独立成分に分離するための非ガウス性独立成分分離処理であることを特徴とする音声分離装置である。非ガウス性独立成分分離処理は他の独立成分処理方法に比べてその前処理としての無相関化処理の影響を受けやすいため、このような第18の発明によれば、最初に非ガウス性独立成分分離処理を行うことにより、無相関化処理がうまく実行されたかどうかを、該無相関化処理に引き続く非ガウス性独立成分分離処理によって効果的に評価することが可能な音声分離装置を実現できる。

【0030】また、本出願に係る第19の発明は、第15乃至第18の発明である音声分離装置において、前記無相関化処理は、少なくとも主成分分析及び因子分析を行うことを特徴とする音声分離装置である。このような第19の発明によれば、各主成分の寄与率を求めて累積寄与率が所定のしきい値を超えるところの成分数を数とすることなどにより、採用する主成分データの数（次数）を決定した上で、効果的に無相関化処理を行うことが可能な音声分離装置を実現できる。

【0031】また、本出願に係る第20の発明は、複数発音者の音声データが混在している混在音声データを、発音者毎の音声データに分離し、該発音者毎の音声データにつき発音者を特定する発音者特定装置において、第15乃至第19のいずれかの発明の音声分離装置により、複数発音者の音声データが混在している混在音声データを、発音者毎の音声データに分離し、分離された前記発音者毎の音声データにつき、発音者毎に該発音者を特定するための特定パラメータを参照して発音者を特定することを特徴とする発音者特定装置である。このような第20の発明によれば、例えば、会議の録音データなどに記録された、複数発音者の音声や雑音などが含まれた混在音声データを音源ごとに分離し、各分離された音声データの発音者を特定することによって、例えば、自動的に会議記録データの作成などを行うことの可能な発音者特定装置が実現できる。

【0032】また、本出願に係る第21の発明は、第20の発明である発音者特定装置において、前記特定パラメータは、発音者が母音を発音した際のホルマント周波数であり、分離された前記発音者毎の音声データにつき、ホルマント周波数を求め、求められたホルマント周波数に關して、前記特定パラメータとしてのホルマント周波数を参照して、発音者を特定することを特徴とする発音者特定装置である。このような第21の発明によれば、フーリエ変換などの容易な処理で抽出できる特徴量であるホルマント周波数を用いて、各分離された音声データの発音者特定を容易に行うことの可能な発音者特定装置が実現できる。

【0033】また、本出願に係る第22の発明は、第21の発明である発音者特定装置において、前記特定パラメータは、発音者が母音を発音した際の第1ホルマント周波数及び第2ホルマント周波数であり、分離された前記発音者毎の音声データにつき、第1ホルマント周波数及び第2ホルマント周波数を求め、求められた第1ホルマント周波数及び第2ホルマント周波数に關して、前記特定パラメータとしての第1ホルマント周波数及び第2ホルマント周波数を参照して、発音者を特定することを特徴とする発音者特定装置である。このような第22の発明によれば、第1と第2のスペクトルピークである2つのホルマント周波数を用いて発音者の特定を行うことにより、容易に、かつより正確に特定を行うことの可能な発音者特定装置が実現できる。

【0034】また、本出願に係る第23の発明は、第20の発明乃至第22の発明のいずれかに記載の発音者特定装置において、分離された前記発音者毎の音声データにつき、前記特定パラメータを参照して発音者を特定できなかった場合には、該音声データから複数の時点のホルマント周波数を求め、求められた複数の時点のホルマント周波数に關して、前記特定パラメータとしての複数の時点のホルマント周波数を参照して、発音者を特定することとを特徴とする発音者特定装置である。このような第23の発明によれば、ある音声の発音者を特定する上の特徴量であるホルマント周波数の、時間的変動をも考慮することにより、より正確に発音者の特定を行うことの可能な発音者特定装置が実現できる。

【0035】また、本出願に係る第24の発明は、第20の発明乃至第23の発明のいずれかに記載の発音者特定装置において、分離された前記発音者毎の音声データにつき、前記特定パラメータを参照して発音者を特定できなかった場合には、該音声データから有声音データを分離し、該有声音データにつき、ホルマント周波数を求め、求められたホルマント周波数に關して、前記特定パラメータとしてのホルマント周波数を参照して、発音者を特定することを特徴とする発音者特定装置である。ホルマント周波数による発音者の特定は、有声音、特に母音の識別に有効であるので、このような第24の発明に

よれば、無声音を含む様々な音声をもより正確に識別することができる。ここで、無相関化処理及び独立成分分離処理がなされる前の音声データが複数人の音声が含まれているデータであるのに対して、無相関化処理及び独立成分分離処理という二つの処理によって分離された分離音声データは、ある一人の音声から抽出されたデータとなっている。よって、このような二つの処理によって分離された分離音声データからは有声音を高い精度で抽出することができる。

【0036】また、本出願に係る第25の発明は、第24の発明の発音者特定装置において、分離された前記発音者毎の音声データにつき、前記特定パラメータを参照して発音者を特定できなかった場合には、該音声データから有声音データを分離し、該有声音データにつき、第1ホルマント周波数及び第2ホルマント周波数を求め、求められた第1ホルマント周波数及び第2ホルマント周波数に關して、前記特定パラメータとしての第1ホルマント周波数及び第2ホルマント周波数を参照して、発音者を特定することを特徴とする発音者特定装置である。このような第25の発明によれば、分離された有声音データに対して、第1と第2のスペクトルピークである2つのホルマント周波数を用いて発音者の特定を行うことにより、より正確に特定を行うことの可能な発音者特定装置が実現できる。

【0037】また、本出願に係る第26の発明は、第24の発明または第25の発明の発音者特定装置において、分離された前記発音者毎の音声データにつき、前記特定パラメータを参照して発音者を特定できなかった場合には、該有声音データにつき、複数の時点のホルマント周波数を求め、求められた複数の時点のホルマント周波数に關して、前記特定パラメータとしての複数の時点のホルマント周波数を参照して、発音者を特定することとを特徴とする発音者特定装置である。このような第26の発明によれば、分離された有声音データに対して、発音者特定上の特徴量であるホルマント周波数の時間的変動をも考慮することにより、より正確に発音者の特定を行うことの可能な発音者特定装置が実現できる。

【0038】また、本出願に係る第27の発明は、第24の発明又は第25の発明の発音者特定装置に關して、前記音声データから前記有声音データを分離する際に、該音声データに対して独立成分分離するための独立成分分離処理が行われることを特徴とする発音者特定装置である。有声音は声帯の振動を伴うものなので、このような第27の発明によれば、音源データに独立成分分離処理をかけることにより、声帯の振動を伴わない無聲音と声帯の振動を伴う有聲音とを容易に分離することの可能な発音者特定装置が実現できる。

【0039】また、本出願に係る第28の発明は、複数発音者の音声データが混在している混在音声データから、議事録を作成する議事録作成装置において、第20

乃至第27のいずれかの発明の発音者特定装置により、分離された前記発音者毎の音声データにつき、発音者を特定し、特定された発音者と、該発音者の発音とを対応付けて記録媒体に出力することにより、議事録を作成することを特徴とする議事録作成装置である。このような第28の発明によれば、発音者の特定が自動的に正確に行われるため、長時間にわたる会議の議事録作成を自動的に行うことの可能な議事録作成装置が実現できる。

【0040】また、第1乃至第5のいずれかの発明の音声分離方法を音声分離装置に実行させるためのコンピュータプログラムも実現可能である。

【0041】また、第6乃至第13のいずれかの発明の発音者特定方法を発音者特定装置に実行させるためのコンピュータプログラムも実現可能である。

【0042】また、そのようなコンピュータプログラムを記録したコンピュータ読み取り可能な記録媒体も実現可能である。

【0043】

【発明の実施の形態】＝混在音声データの音声分離＝

以下、図面を参照しつつ、本発明のより具体的な実施形態につき、詳細に説明する。まず、本発明の方法の前半部分である、混在音声データの音声分離ステップについて説明する。

【0044】本実施形態では、2人で行われたる会議の発言内容の音声データを2本のマイク（マイク1、マイク2）で拾う。図1は、そのうちマイク1から入力された音声データ（生データ）Xの波形である。この混在音声データには、複数の発音者の音声データが混在しているのみならず、音楽や、さらには雑音などが混ざっているもよい。2人の発音をそれぞれ音源S1、S2と呼ぶことにする。

【0045】図2は、音声分離処理のサイクルを示す図である。マイク1及びマイク2から入力された混在音声データは、まず無相関化処理W1にかけられる。無相関化処理W1に渡される音声データは、図1の[1]、[2]のようにセグメント化されて1つずつ渡される。最も効率がよいように、各セグメントは互いに1/2周期ずつオーバーラップしている。

【0046】図2において、無相関化処理W1の次のステップである1Cチューナーは、独立成分解析（ICA）の手法を3種類のうちから選択するためのチューナーである。その次のステップである独立成分分離処理W2は、非ガウス性に基づく分離処理W2（α）、非定常性に基づく分離処理W2（β）、有色性に基づく分離処理W（γ）の3種類のうちいずれかの方式の処理を行う。W2の後のステップの評価器Eでは、W2にて分離されたデータの分離性の評価を行う。マイクから入力された混在音声データの音声分離性能が充分になるまで、以上のW1→1Cチューナー→W2→Eというサイクル

を繰り返す。ただし、1回目のサイクルでは、独立成分分離処理W2として、非ガウス性に基づく独立成分分離処理W2（α）を行い、2回目以降のサイクルでは、1Cチューナーの選択に従って、W2（α）、W2（β）、W2（γ）の3種類のうちから適切な方式の独立成分分離処理を行う。

【0047】図3は、1回目の音声分離サイクルを示している。図1における前記[1]の時間セグメントの、マイク1及びマイク2からの混在音声データx1、x2

が、まず無相関化処理W1に入力される。

【0048】図7及び図8は、それぞれx1及びx2のデジタル化波形図データ（縦軸は音の強さで、単位はミリボルト）を示す。各時点のx1、x2データを、横軸をx1の強さ、縦軸をx2の強さとして散布図を描くと図9のようになる。散布図は、第1象限から第3象限にかけて若干直線的な分布を呈し、x1とx2のデータは互いに相関性を有することを示している。これら生データであるx1、x2が無相関化処理W1にかけられると、互いに相関性を有しないデータf1、f2に変換される。

【0049】f1及びf2の散布図を図10に示す。図10の横軸は因子得点Fの第1因子f1、縦軸は因子得点Fの第2因子f2を示している。図9が軸に対してほぼ平行四辺形状に分布していたのに対し、軸に対してほぼ十字形で分布した形状に分布しており、f1とf2とは互いに相関性を有していないことがわかる。

【0050】ここで、無相関化処理の内容について説明する。図6は、無相関化処理W1の一例のフローチャートを示したものである。まず、図7及び図8に示した音声生データx1、x2を（1）式により標準化する。標準化の結果、平均が0、標準偏差1のデータとなる。

【数1】

$$X = \frac{(\bar{X} - \mu)}{\sigma} \quad (1)$$

【0051】生データx1、x2の相関行列（ベクトルC）を（2）式より求める。（2）式において（x1、x2）はベクトルの内積を表す。

【数2】

$$C = \frac{1}{n} (X_1, X_2) \quad (2)$$

【0052】上記相関行列に対する固有値λ1と固有ベクトルAを（3）式より求める。

【0053】

$$CA = \lambda A \quad (3)$$

【0054】今、因子分析によって、互いに無相関な因子得点を求めようとしているのだが、その際、第1番目の因子から始めて、何番目の因子までを採用するのか、重要な点である。m番目の因子までを採用する場合を、

m次元と呼ぶ。先に求めた固有ベクトルAにより、
式(4)によって主成分Zが求まる。

【数4】

$$Z = A \cdot X \quad (4)$$

【0055】次にm側の因子に対して、(5)式の形の
近似式にて因子分析を実行する。(5)式におけるe
は、特殊因子と呼ばれるものである。

【数5】

$$X = A \cdot Z + e \quad (5)$$

【0056】この因子モデルが(6)式の表現をとる。
(6)式における因子負荷量b_{ij}、因子得点Fは、
(7)式及び(8)式によって求める。そして、図6の
フローチャートの最終ステップで、結局音声生データ
は、互いに無相関な因子得点(ベクトルF)に変換され
る。

【数6】

$$X = b \cdot F + e \quad (6)$$

【数7】

$$b_{ij} = \sqrt{\lambda_j} A_{ij} \quad (7)$$

【数8】

$$F = (b^T b)^{-1} b^T X \quad (8)$$

【0057】以上説明したW1の主な特徴は、主成分分
析と因子分析とを組み合わせている点である。その効果
は、主成分分析を実行すると各主成分の寄与率を同時に
求めることができるので、例えば、第1次主成分から第
m次主成分までの累積寄与率が80%を超えるまでの主
成分を採用するようにすることで、次数mを決定するこ
とにある。分離すべき音声生データは、時間的変動が大
きく、混合による相関の度合いが大きく変化するので、
何個の因子を採用するかは無相関化処理において重要な
点である。

【0058】発話者の人数があらかじめ判明している場
合には、次数mを発話者の人数に固定してしまえばよい
が、人数が不明なときは、例えば、累積寄与率が所定
のしきい値を超えたときの主成分数を次数mとする。次数
mの決定方法は、システムに応じて様々な方法を準備し
ておき、臨機応変に変化させる(チューニングする)こと
が好ましい。次にこのチューニングの実施例について詳
しく説明する。

【0059】図23は、システムに応じた方法で次数m
を決定する手順を示すフローチャートである。図23
で、R・Kは累積寄与率の初期しきい値、Mは採用し得
る最大次数(次数の上側しきい値)、△R・Kは累積寄与
率の変化量である。主成分分析を実行すると、図21の
ような、次数m(第m主成分まで採用した)ということ
を示す。図23にはA、B、C3種類のグラフの例を描いて
いる。

【0060】まず、第1の処理ステップとして、累積寄

与率R・Kにしきい値R・K0(この実施例では80%)を
設定しておき、このしきい値R・K0を超える次数mを求
める。ところが、次数があまりに大きくなりその後の処理
が煩雑に過ぎるので、あらかじめ次数の上限値Mを決め
ておく。図21の例では、M=4とすると、Aの場合はし
きい値R・K0を超える次数m=2であるので、m=2
< 4=Mとなっており、次数mは2に決定される。Bの例
ではR・K0を超える次数mは5であるので、m=5>4=M
となってしまう、次数mはまだ決定されない。Cの例
でも同様に次数mは決定されない。

【0061】そのような場合は図22に示す、第2のステ
ップを実行する。すなわち、次数mの増加に対する、
R・Kの差分変化量△R・Kを調べる。これは要するに、累
積寄与率の変化が最大となる次数mをもって採用すべき
次数とするという処理方法である。この実施例では、B
の例ではm=2、Cの例ではm=4において△R・Kが最大
値をとる。この場合も次数mが上限値Mよりも下なら
ば、その次数mを採用とするが、Mを上回る場合は、そ
の処理が次のステップに送られる。

【0062】第2のステップでも次数mが上限値Mを超
えてしまう場合であれば、次に累積寄与率のしきい値R
・K0を引き下げて、例えば60%(=R・K1)とし、上
記第1のステップと同じように比較する。新しいしきい
値R・K1を超えるところの次数がM=4以下であれば、
これを次数mとして採用とし、Mを超える場合は、所定
の下げ幅で順次R・K2、R・K3、・・・R・Knの値を下
げる。ただし、累積寄与率R・Kが50%を下回るという
ことは、半分以上の情報が失われてしまうことを意味す
るので、R・Knの下限値は50%とする。

【0063】次数mがR・Kn=50%以上で、かつM以
下の値で発見されない場合は、再び上記第2のステップ
と同様の処理、すなわち△R・Kが最大になる次数を求め
て、その値を次数mとして採用してしまう。これは、累
積寄与率が大きく変化するということは、その次数の前
後で情報がより多く保存されるということを意味するの
で、少なくともその次数までは採用したい、という考え
に基づくものである。

【0064】以上のようにして、図3において、無相関
化されたデータf1、f2は、あたりに独立成分分離処理
W2に送られる。1回目の音声分離サイクルでは、こ
れらの無相関化データf1、f2に対し、非ガウス性
に基づく独立成分分離処理W2(a)を実行する。

【0065】以上、図3におけるW1及びW2(a)の処
理により、分離信号aおよびbが得られ、これらの分離
性(充分に分離されているか否か)を評価器Eで評価
し、分離が不十分なとき(図の※1)はこれら、a、bの
データに対して、2回目のサイクルを実行する。

【0066】2回目のサイクルの例を図4に示す。図3
に示した1回目のサイクルと似ているが、1Cチューナ
ーにおける処理が加わっている。独立成分分離処理W2

を行う前に、1 Cチューナーで2回目の無相関化処理されたデータ「1」、 「2」の信号特性を解析し、非ガウス性に基づく処理W2 (a)、非定常性に基づく処理W2 (b)、有色性に基づく処理W2 (c)のいずれをW2として実行するかを選択する。この例ではW2 (b)を実

された入力データのいずれもが有色性を有すると評価された場合、1 Cチューナーは、独立成分分離処理W2として有色性に基づく処理W2 (c)を選択する。

【0071】図5は3回目のサイクルを示している。各処理は2回目のサイクルと同様であるが、3回目の独立